DOI: https://doi.org/10.37547/supsci-oje-05-03-02 **Pages: 9-19**

Oriental Journal of Education

ORIENTAL JOURNAL OF EDUCATION

journal homepage:
https://www.supportscience.uz/index.php/oje

A STUDY ON THE ISSUES IN USING KOREAN LANGUAGE LEARNER CORPORA

Choi Jeong Doh
Associate Professor
College of Humanities and International Studies, Keimyung University
Korea
E-mail: uncleshadow@kmu.ac.kr

ABOUT ARTICLE

Key words: Korean learner's corpus, linguistic research, corpus utilization, corpus characteristics, corpus construction process, online search tool.

Received: 26.04.25
Accepted: 28.04.25
Published: 01.05.25

Abstract: Language data produced by learners of foreign languages is called a *learner's corpus*. Among these, if the target of learning is Korean, it is referred to as a *Korean learner's corpus*. A notable feature of the Korean learner's corpus is that it is developed and provided by the government. This corpus is made available in its entirety with resolved copyright issues, and it also offers a web-based search tool that supports researchers. However, if there is no *understanding* of how the corpus was constructed, there is a critical risk that the research outcomes may lack credibility. Therefore, this study outlines precautions when using the Korean learner's corpus. First, it is necessary to check the *guidelines* in the published report from when the corpus was created; second, it is important to identify *features* that are not stated in the report by using the *search function* of the corpus; and third, it is emphasized that users must make personal efforts to understand the characteristics of the corpus they are using.

KOREYS TILI O'RGANUVCHILARINING KORPUSDAN FOYDALANISH MUAMMOLARI TADQIQI

Choi Jeong Doh
professor
Koreya Respublikasi Keimyung universiteti
Koreya
E-mail: uncleshadow@kmu.ac.kr

MAQOLA HAQIDA

Kalit so'zlar: Koreys tili o'quvchi korpusi,

Annotatsiya. Chet tillarini o'rganuvchi

tilshunoslik tadqiqoti, korpusdan foydalanish, Korpus xususiyatlari, korpus yaratish jarayoni, onlayn izlash vositasi.

o‘quvchilar tomonidan yaratilgan til ma'lumotlari “o‘quvchi korpusi” (learner’s corpus) deb ataladi. Ulardan o‘rganish maqsadi koreys tili bo‘lganlari “koreys tili o‘quvchi korpusi” deyiladi. Koreys tili o‘quvchi korpusining xususiyati shundaki, u davlat (hukumat) tomonidan yaratilgan va taqdim etilgan. Ushbu korpus mualliflik huquqlari hal qilingan holda to‘liq ma’lumotlarni taqdim etadi va veb orqali izlash vositasi (Search Tool) bilan ta’minlangan bo‘lib, tadqiqotchilarga yordam beradi. Biroq, agar korpus qanday yaratilgani haqida “tushuncha” bo‘lmasa, tadqiqot natijalarining ishonchliligiga jiddiy shubhalar tug‘ilishi mumkin. Shu sababli, ushbu tadqiqotda koreys tili o‘quvchi korpusidan foydalanishda e’tibor berilishi lozim bo‘lgan jihatlar ko‘rsatib o‘tilgan. Birinchidan, korpus yaratilganida chop etilgan hisobotdagi “yo‘riqnomalar”ni tekshirish zarur; ikkinchidan, korpusni “izlash” orqali hisobotda ko‘rsatilmagan "xususiyatlar"ni aniqlash kerak; uchinchidan, foydalanuvchining korpus xususiyatlarini tushunishga bo‘lgan shaxsiy harakati muhimligi ta’kidlangan.

한국어 학습자 말뭉치 사용에서의 쟁점 연구

최정도
계명대학교 인문국제학대학 한국어교육과
조교수
010-4744-8152
uncleshadow@kmu.ac.kr

О СТАТЬЕ

키워드: · 한국어 학습자 말뭉치, 언어 연구, 말뭉치 활용, 말뭉치 특성, 말뭉치 구축 과정, 온라인 검색 도구

초록. 외국어를 배우는 학습자들이 생산한 언어데이터를 학습자말뭉치(learner’s corpus)라고 한다. 그중에서 학습의 목표가 한국어인 것을 한국어학습자말뭉치라고 하는데, 한국어학습자말뭉치는 ‘국가(정부)’에서 만들어서 제공하고 있다는 특징이 있다. 이 말뭉치는 저작권이 해결하여 전체 데이터를 제공하고 있으며, 웹으로 검색 도구(Search Tool)을 제공하고 있어 연구자들에게 도움을 주고 있다. 그러나 말뭉치가 어떻게 만들어졌는지에 대한 ‘이해’가 없다면, 연구 결과가 신뢰를 받지 못하는 치명적인 문자가 발생할 수도 있다. 따라서 이 연구에서는 한국어학습자말뭉치를 활용할 때의 유의점에 대해서 제시하였다. 첫째, 말뭉치를 만들 때 발간된 보고서의

'지침' 확인이 필요하고, 둘째 말뭉치 '검색'을 통해서 보고서의 나타나 있지 않은 '특징'을 확인하는 것이 필요하며, 셋째 활용 말뭉치의 특성을 파악하기 위한 사용자 스스로의 노력이 필요하다는 것을 언급하였다.

1. 서론

이 연구에서는 대학원 수업 시간이나, 학위 논문 작성을 앞둔 학생들과 면담(지도)에서 빈번히 나타날 수 있는 한국어 학습자 말뭉치 사용에서의 몇 가지 기초적인 사항을 점검하고자 한다.

지금 현재는 한국에서 '21세기 세종계획; 세종 말뭉치'라고 하는 대규모 말뭉치가 연구용 목적으로 공개된 이후, 언어 연구에서 말뭉치의 사용은 특별한 일이 아닌 현실이 되었다. 연구용 말뭉치의 공개와 사용 도구(글잡이, 한마루, 앤트콘크/지능형형태소분석기, 유태거, 키위 등등)의 공개로 학생들의 능력 또한 과거의 달리 월등히 늘어난 것이 현실이다. 그리고 AI의 발전과 '모두의 말뭉치'의 공개로 더욱 다양한 유형의 언어 자료를 사용할 수 있는 환경도 구축되었다.

그러나 이러한 환경의 발전에도 불구하고 '장면1'과 '장면2'는 대학원 수업 시간이나, 학위 논문 작성을 앞둔 학생들과 면담(지도)에서 빈번히 나타날 수 있는 상황인 것이 또 피할 수 없는 현실로 생각된다. 연구의 주제, 방향, 내용을 지도하기도 벅찬 상황에 말뭉치의 소개, 말뭉치의 특징 소개, 말뭉치의 사용 방법 안내까지 해야 하는 상황이 발생할 수 있을 것이다.

이 논의는 말뭉치와 말뭉치 활용 도구의 특성을 잘 이해하지 못한다면, 잘못된 결과가 도출된다는 사실에 관심을 가지고 '한국어 학습자 말뭉치'를 중심으로, 몇 가지 사용에서의 유의점에 대해서 언급하는 것이 목적이다.

2. 한국어 학습자 말뭉치란?

여기서는 먼저 한국어 학습자 말뭉치의 재원에 대해 개략적인 사항에 대해서 살펴보고자 한다. 이미 잘 알려져 있듯이 '한국어 학습자 말뭉치'는 학습자가 생산한 한국어 언어 자료로 국립국어원에서 2015부터 꾸준히 구축하여 배포하고 있는 자료이다. 현재 2024년 기준 146개국, 101개 언어권의 자료가 사용역별, 국적별, 언어권별, 수준별, 나이대, 내용별, 목적 등을 고려하여 구축되었다. 개략적인 사항은 다음과 같다.

※ 한국어 학습자 말뭉치 -

- 학습자가 생산한 말뭉치: 국립국어원 구축(2015~)
- 146개국, 101개 언어권(2024.07. 현재)

- 사용역별, 국적별, 언어권별, 수준별, 나이대, 내용별, 목적 등

사용역별: 문어, 구어

국적별: 중국, 일본, 베트남, 홍콩, 타이, 대만, 기타

언어권별: 중국어, 일본어, 베트남어, 영어, 중국어(광둥어), 러시아어, 기타

나이대: 10대, 20대, 30대, 40대, 50대 이상

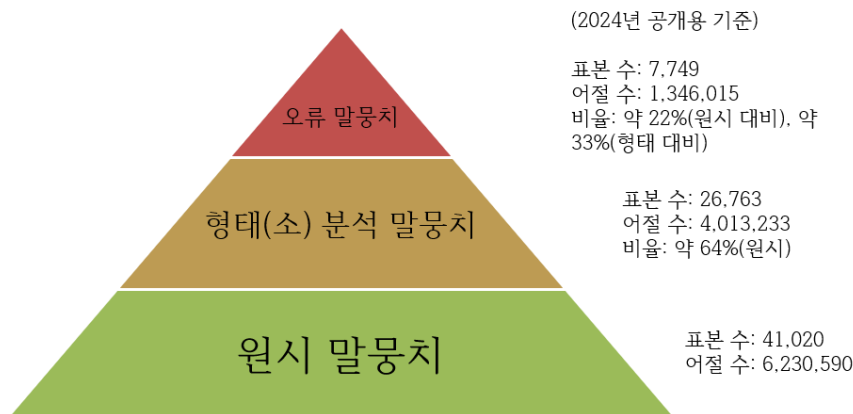
급수별: 1급, 2급, 3급, 4급, 5급, 6급, 6급 이상 등

그리고 '한국어 학습자 말뭉치'는 분석의 수준(층위: 부가되는 언어 정보의 층위)에 따라서 원시 말뭉치와 형태(소) 분석 말뭉치, 오류 (분석) 말뭉치로 계층적으로 구축되어 있다.

(분석 수준에 따른) 말뭉치의 종류

- 원시 말뭉치: 언어 자원을 있는 그대로 입력해 놓은 말뭉치
- 형태(소) 분석 말뭉치: '원시 말뭉치'에 어절별로 형태 분석을 한 말뭉치
- 오류 (분석) 말뭉치:

(좁은 의미) 학습자의 오류를 분석한 말뭉치

(넓은 의미) 학습자가 생산한 말뭉치(오류가 포함되어 있다는 의미에서)

3. 한국어 학습자 말뭉치의 활용 방법

'한국어 학습자 말뭉치'를 활용할 때에는 대개 분석 수준에 따른 종류인 원시 말뭉치, 형태(소) 분석 말뭉치, 오류 (분석) 말뭉치를 대상으로 한다.

먼저 원시 말뭉치의 활용의 경우 검색어를 통한 용례의 검색 가능하고 많은 양과 표본에서 검색이 가능하다는 장점이 있다. 아울러 국립국어원에서 이 말뭉치를 활용할 수 있는 웹검색기를 제공하고 있는데, '그리고, 또는' 검색과 다양한 조건(급수, 언어권 등)을 선택할 수 있는 '상세 검색'을 지원하고 있어 사용자 입장에서 편리하다.

그런데 원시 말뭉치를 대상으로 연구를 진행할 때의 검색은 음절의 수가 길거나, 검색 형태가 다른 형태와 겹치지 않는 검색어일 경우에 효용이 크다. 주로 '명사, 부사, 감탄사' 등 이

나 '조사, 어미, 접사' 중 음절 수가 길거나 형태가 독특한 것이 대상이 된다. 불변어일 경우에는 원시 말뭉치 활용의 효용을 가질 수 있으나 용언은 활용형 전체를 검색하는 것은 현실적으로 불가능하기 때문에(만약 원시를 이용한다면 개별로 모두 검색해야 함.) 추천되지 않는다.(예: 먹은, 먹는, 먹을, 먹었다, 먹을까, 먹었습니다, 먹어요 등) 그리고 단음절을 연구 대상으로 삼아 원시 말뭉치를 검색하는 것은 추천되지 않는다. 학습자의 오류 역시 찾기 어렵다. 대안으로는 형태-오류 주석 말뭉치 사용하는 것이 추천된다.

– 한국어 학습자 말뭉치의 활용 –

- **원시 말뭉치의 활용**
  - 검색어를 통한 용례의 검색 가능**(장점: 많은 양과 표본에서 검색이 가능함.)**
  - '그리고, 또는' 검색 지원
  - 다양한 조건(급수, 언어권 등)을 선택할 수 있는 '상세 검색' 지원

- **원시 말뭉치 검색의 팁**
  - 음절의 수가 길거나, 검색 형태가 다른 형태와 겹치지 않는 검색어 추천
    – 주로 '명사, 부사, 감탄사' 등 이나 '조사, 어미, 접사' 중 음절 수가 길거나 형태가 독특한 것
  - 용언은 활용형 전체가 검색되지 않음.(만약 원시를 이용한다면 개별로 모두 검색해야 함.)
    – 예) 먹은, 먹는, 먹을, 먹었다, 먹을까, 먹었습니다, 먹어요 등
  - 학습자의 오류는 찾기 어려움 ⇨ **(대안) 형태-오류 주석 말뭉치 사용**

4. 한국어 학습자 말뭉치 활용의 쟁점

여기서는 한국어 학습자 말뭉치를 어떠한 목적으로든 활용하고자 할 때, 연구의 결과가 왜곡되지 않도록 꼭 숙지하고 있어야 하는 것들에 대해서 언급하고자 한다.

4.1. 말뭉치와 활용 도구의 확보

다음 교실에서의 '장면1'을 구체적으로 상정한 후, 그에 대해서 살펴보자.

장면1-1: "교수님! 말뭉치는 어디에 있어요? 어디서 받을 수 있어요?"

말뭉치를 연구자 자신이 어떠한 연구를 하는지, 어떠한 말뭉치를 사용하는지에 따라서 활용해야 할 말뭉치가 다르다. 대략적으로 한국어 언어 자료는 대부분 '국립국어원' 홈페이지에서 해소가 가능하다. 국립국어원에서는 '세종 말뭉치', '모두의 말뭉치', '한국어 학습자 말뭉치', 사전 데이터(표준국어대사전, 우리말샘, 한국어기초사전) 등의 대규모의 다양한 언어 자료를 제공하고 있어 연구자들에게 도움을 준다.

장면1-2: "그런데, 어떻게 봐요? 어떻게 사용해요?"

말뭉치는 뷰어로서의 워드프로세서나 웹창에서도 확인이 가능하다. 하지만 데이터를 가공하거나 변환을 해야 하기 때문에, 대부분 무료로 사용이 가능한 텍스트 에디터(UltraEdit, Emeditor, Editplus 등)의 사용을 권장하는 편이다. 한편 말뭉치를 검색하는 도구는 생각보다 많지 않다. '학습자 말뭉치'의 경우는 국립국어원의 '한국어 학습자 말뭉치 나눔터'를 활용할 것

을 추천한다.

- 세종 말뭉치: 한마루 등

- 모두의 말뭉치: 없음(곧 웹 서비스 개시할 것으로 보임.)

4.2. 한국어 학습자 말뭉치 활용 시의 유의점

여기서는 '장면2'에서 말뭉치 확보 이후 활용하는 데서 나타날 수 있는 유의점에 대해서 유형별로 살펴보고자 한다.

장면2-1: "교수님, '공부하다'가 말뭉치에 없어요.", "교수님, '적극적'이 말뭉치에 없어요."

장면2-2: "교수님, 접미사 '-히, -이'가 말뭉치에 없어요."

"교수님, 피동사하고 사동사는 어떻게 찾아요?"

4.2.1. '있다'가 없다?

'한국어 학습자 말뭉치'는 '21세기 세종 말뭉치'의 구축 방식을 토대로 구축되었다. '세종 말뭉치'는 구축 시 형태 분석 말뭉치 구축의 어려움으로 인하여 '있다'를 동사(VV)로 통합하여 분석하였는데, 이러한 양상이 '한국어 학습자 말뭉치'에도 그대로 이어지고, 분석기에도 적용되어 있다. 다만, '모두의 말뭉치'에서는 이러한 문제점을 보완하기 위하여 '있다'의 형용사(VA) 분석이 적용되어 있다.

- 2016\(7574)홍콩_광둥어_2급_20_취업;취미_문어_국내\sample_7574.xml

시간이 있을 때 공연을 보는 것을 좋아해요. **있을있/VV + 을/ETM**

- 2016\(7574)홍콩_광둥어_2급_20_취업;취미_문어_국내\sample_7574.xml

한국에서 홍콩 음식을 먹을 수 있으면 정말 행복해요.**있으면있/VV + 으면/EC**

4.2.2. 접사 분석의 방법

말뭉치를 활용할 때에는 반드시 '말뭉치 구축 지침'을 확인해야 할 필요가 있다. 한국어 말뭉치는 한국어학의 성과와 한국어 사전 정보를 반영하고 있지만, 구축의 어려운 현실로 인하여 항상 내가 알고 있는 정보가 그대로 반영되어 있는 것은 아니다. 그렇기 때문에 반드시 지침을 확인해야 한다. 이러한 말뭉치의 구축 지침은 국립국어원 홈페이지(연구 보고서)에서 제공하고 있으니, 말뭉치를 활용하는 이들은 지침을 통하여 말뭉치의 구축 현황을 숙ㅈ고할 필요가 있겠다.(세종 말뭉치, 모두의 말뭉치, 한국어 학습자 말뭉치)

예를 들어 '접사' 연구에서 이들 말뭉치를 활용할 경우 전체 목록이나 개별 목록에서 문제가 발생하는 경우가 있다. 따라서 말뭉치 구축 지침에서 분석된 '접사 목록' 확인 필요하다.

- 국립국어원(2020:124-128)의 접사 목록(세종말뭉치의 업그레이드)

명사 파생 접두사 (XPN): 33개	가(假), 고(高), 과(過), 구(舊), 날, 노(老), 대(大), 맏, 맨, 무(無), 미(未), 반(反), 범(汎), 부(不), 불(不), 비(非), 생(生), 소(小), 신(新), 왕(王), 재(再), 저(低), 제(第), 준(準), 초(超), 최(最), 친(親), 탈(脫), 폐(廢), 풋, 피(被), 한, 헛
명사 파생 접미사 (XSN): 64개	가(哥), 가(價), 가량, 간(間), 경(頃), 계(界), 계(系), 광(狂), 권(券), 권(圈), 권(權), 기(氣), 께, 꼴, 꾼, 끼리, 네, 님, 당(當), 대(臺), 댁(宅), 들, 들이, 론(論), 류(類), 률(率), 리(裡), 발(發), 배기, 별(別), 부(附), 분(分), 분지(分之), 뻘, 산(産), 상(上), 생1(生), 생2(生), 성(性), 시(視), 씩, 어치, 여(餘), 용(用), 율(率), 장이, 쟁이, 적(的), 정(整), 제(制), 질, 짜리, 째1, 째2, 쯤, 층(層), 치(値), 치레, 투성이, 풍(風), 하(下), 형(型), 형(形), 화(化)
동사 파생 접미사 (XSV): 5개	당하(다), 되(다), 시키(다), 하(다), 받(다)
형용사 파생 접미사 (XSA): 5개	답(다), 되(다), 롭(다), 스럽(다), 하(다)

'21세기 세종 말뭉치'와 '한국어 학습자 말뭉치'는 위 표에 있는 접사가 포함된 단어의 경우 모두 분석하게 된다. 즉 위 표에 없는 접사의 경우는 분석하지 않다는 것이다. 따라서, '파생어' 연구나, '접사' 연구를 위해서는 연구자 자신이 말뭉치에 손을 대거나, 어휘 빈도의 결과를 두고 분석해야 한다.

이들의 특징은 대체적으로 생산적인 접사만 분석의 대상으로 삼았다는 것이다. 따라서 접사 연구자들은 연구 대상이 되는 접사가 말뭉치에 분석이 되어 있는지 반드시 확인을 해야 하다. 앞선 '피동사'와 '사동사'에 대한 장면에서 이들 말뭉치는 '-이/히/리/기/우/구/추-' 등의 파사동 접사는 분석하지 않았다는 사실을 알아둘 필요가 있겠다.

- 공부/NNG + 하/XSV / 적극/NNG + 적/XSN / 조용/XR + 하/XSA

이를 통해 보자면 '적극적'과 같은 파생명사는 원시 말뭉치에서 검색하는 것이 더 효과적일 수 있으나, 파생용언인 '공부하다' 등의 검색은 원시 말뭉치 사용이 상당히 제한된다는 것을 알 수 있다. 아울러 한국어 말뭉치에서는 '합성어'는 더 이상 분석하지 않는다는 사실도 알아둘 필요가 있겠다.

- 접사 분석 표지(태그, tag)

범주	표지(태그, tag)	비고
명사 파생 접두사	XPN	
명사 파생 접미사	XSN	
동사 파생 접미사	XSV	

형용사 파생 접미사	XSA	
동사 파생 접두사	XPV	추가(되찾다, 휘감다)
형용사 파생 접두사	XPA	추가(새까맣다)
부사 파생 접미사	XSM	추가(조용히, 깨끗이)

하__03/XSV 171624			
하__03/XSA 62529	음__10/XSN 1418	별__04/XSN 494	국__10/XSN 201
이__32/XSM 49235	장__45/XSN 1533	비__32/XPN 466	히__07##이/XSM196
들__09/XSN 48661	톱/XSA 2288	우__18/XSV 455	지__10/XSA 188
어__08/XSN 16199	성__17/XSN 2282	물__09/XSN 402	신__15/XPN 188
ㅁ__03/XSN 15691	점__11/XSN 1793	사__37/XSN 385	드리/XSV 187
님__04/XSN 13473	관__19/XSN 1777	학__06/XSN 374	증__11/XSN 324
적__18/XSN 12693	기__38/XSN 1175	회__14/XSN 611	화__88/XSN 303
히__07/XSM 8569	력__02/XSN 1150	시__27/XPN 606	증__12/XSN 293
이__31/XSV 7599	이__29/XSN 1131	비__33/XSN 603	객__02/XSN 272
되__05/XSV 5960	품__08/XSN 1057	무__11/XPN 583	추__09/XSV 267
인__17/XSN 5518	시키__02/XSV 957	율__04/XSN 582	기__37/XSN 261
대__17/XPN 5442	여__26/XPN 932	찍__03/XSN 569	탕__08/XSN 252
일__09/XSN 4046	히__06/XSV 907	지__26/XSN 548	제__19/XSN 241
리__18/XSV 3854	원__17/XSN 896	다__09/XPN 518	외__08/XPN 239
째__02/XSN 3199	스럽/XSA 886	저__11/XPN 367	어__08##꺼/XSN 235
자__31/XSN 2959	실__12/XSN 832	사__42/XSN 364	제__21/XPN 233
화__16/XSN 2630	기__37/XSV 807	형__08/XSN 364	튤__04/XSN 232
이우__08/XSV 2379	관__20/XSN 792	금__14/XSN 354	국__09/XSN 222
생__08/XSN 1706	식__09/XSN 747	원__19/XSN 341	법__03/XSN 216
쯤__02/XSN 1628	남__09/XPN 696	심__09/XSN 329	되__05/XSA 213
원__18/XSN 1611	불__15/XPN 612	민__07/XSN 204	사__41/XSN 209
기__43/XSN 1280	가__13/XSN 499	어__08##거/XSN 203	끼리/XSN 206
갈__18/XSN 1236			

4.2.3. 수사와 수관형사의 미분리

'한국어 학습자 말뭉치'는 수사와 수관형사의 형태가 같을 경우 이들을 구분하지 않고 모두 수사(NR)로 분석하고 있다. 따라서 말뭉치를 활용하여 수관형사를 연구할 때에는 상당히 유의할 필요가 있다.

2015\(8174)일본_일본어_6급_10_취미_구어_국내\sample_8174.xml

영화는 그 너무너무 옛날의 이야기 아니라 십 년 전 십**십/NR** 년 년/NNB

4.2.4. 인용 표지의 분석

국립국어원의 모든 말뭉치는 인용 표지의 경우 따옴표가 있을 때와 없을 때를 구분하여 분석하고 있다는 사실도 알아둘 필요가 있다. 따라서 말뭉치를 활용하여 한국어 인용을 연구할 경우에는 첫째, 인용표지가 나타나는 경우와 나타나지 않는 경우를 모두 고려해야 하며, 둘째, 인용 표지가 나타날 경우 말뭉치에서는 인용 표지가 없을 경우 연결어미로 분석했다는 사실을 꼭 알아둘 필요가 있겠다.

- 직접 인용

2015\(1197)필리핀_타갈로그어_2급_10_진학_문어_국내\sample_1197.xml

~ 내 전공은 공학입니다"라고 얘기해서 다른 말에 대한 이야기로 하려고 했다.

공학/NNG + 이/VCP + ㅂ니다/EF + "/SS + **라고/JKQ**

- 간접 인용

2015\(1193)태국_타이어_2급_20_취미_문어_국내\sample_1193.xml

그리고 나에게 고양이를 키울 수 있지만 비밀이라고 했다.

비밀/NNG + 이/VCP + **라고/EC**

4.2.5. 기능어 연구 시 유의점

한국어 분석 말뭉치는 (기능어의 경우) 이형태를 반영하고 있다. 그렇기 때문에 말뭉치를 활용한 기능어 연구에서는 분석에서 적용된 이형태의 목록을 확인할 필요가 있다.

- '-었-/-았-', '-ㄴ/은', '-다/ㄴ다/는다', 'ㄹ/을' 등

문장 형성소(고영근)의 경우, 모두 분석적 차원에서 구분을 하고 있으나 문장 구성소(고영근)의 경우, 통합형으로 분석하는 것이 한국어 분석 말뭉치의 특징이다.

- 관형사형 어미(ETM):

ㄴ/ETM, 는/ETM, ㄹ/ETM, 은/ETM, 던/ETM, 을/ETM, 다는/ETM, 라는/ETM, ㄴ다는/ETM, 란/ETM, 려는/ETM, 자는/ETM, 는다는/ETM, 단/ETM, 느냐는/ETM, 으려는/ETM, 냐는/ETM, ㄴ단/ETM, 리라는/ETM, 으리라는/ETM, 려던/ETM, 더라는/ETM, 랄/ETM, 으라는/ETM, 다던/ETM, 넌/ETM, 더란/ETM, 냔/ETM, 라던/ETM, 는단/ETM, ㄴ가라는/ETM, 으란/ETM, 다라는/ETM, 으리란/ETM, 으려던/ETM, 리란/ETM, 잘/ETM, 노라는/ETM, ㄴ다던/ETM, 고픈/ETM, 는가라는/ETM, 느냔/ETM, 자던/ETM, 잔/ETM, 달/ETM, 으냐는/ETM, 아란/ETM, 리라던/ETM, 너냔/ETM, 능/ETM, ㄴ다라는/ETM, ㄹ까라는/ETM, 아서라는/ETM, 다넌/ETM

- 명사형 어미(ETN):

기/ETN, ㅁ/ETN, 음/ETN, 라기/ETN, 다기/ETN, ㄴ다기/ETN, 이/ETN, 는다기/ETN

- '-더' 통합형(종결 어미):

더라/EF, 더군/EF, 더라구/EF, 더란다/EF, 더냐/EF, 다더라/EF, 더구나/EF, 더라고/EF, 더니/EF, 더라니까/EF, 더래/EF, ㄴ다더라/EF, 다더군/EF, 더구만/EF, 더랍니다/EF, 더이다/EF, 라더라/EF, 더구먼/EF, 더만/EF, 더랬다/EF, 더라도/EF, 라더냐/EF, 더노/EF, 더라니/EF, ㄴ다더냐/EF, 더라야/EF, 더라나/EF, 더니만/EF

- '-더' 통합형(연결 어미):

더니/EC, 더라도/EC, 더라면/EC, 더라/EC, 더라고/EC, 더니만/EC, ㄴ다더니/EC, 라더니/EC, 다더니/EC, 더면/EC, 더냐/EC, 더냐고/EC, 더군요/EC, 다더라/EC, ㄴ다더라/EC, 라더라/EC, 더래두/EC, 더라구/EC, 더라면서/EC, 더라며/EC, 더라니/EC, 더라두/EC, 더마는/EC, 더라구요/EC, 더구먼/EC, 는다더니/EC

- '요' 결합하는 종결 어미: '요' 통합형으로 분석

어요/EF, **지요/EF**, **아요/EF**, ㄹ까요/EF, 군요/EF, 거든요/EF, 나요/EF, 더군요/EF, 네요/EF, 잖아요/EF, ㄴ가요/EF, 을까요/EF, 니까요/EF, 으니까요/EF, 구요/EF, 는데요/EF, ㄴ데요/EF, 세요/EF, 대요/EF, 고요/EF, ㄹ게요/EF, 는군요/EF, 예요/EF, 다구요/EF, 라구요/EF, 더라구요/EF, 던가요/EF, ㄴ대요/EF, 어야지요/EF, 래요/EF, ㅂ니다요/EF, 라고요/EF, 던데요/EF, 아야지요/EF, 는지요/EF, 어서요/EF, 은데요/EF, 다니까요/EF, 라니까요/EF, 아서요/EF, ㄹ께요/EF, 습니다요/EF, 더라고요/EF, 다니요/EF, 라니요/EF, ㄹ래요/EF, ㄴ다구요/EF, 게요/EF, 습지요/EF, 지만요/EF, ㄴ지요/EF, ㅂ지요/EF, 다고요/EF, 냐구요/EF, 는걸요/EF, 다면서요/EF, ㄴ다니까요/EF, 구만요/EF, 라요/EF, 로군요/EF, 야요/EF, 더라니까요/EF, 습니까요/EF, 려구요/EF, ㅂ니까요/EF, ㄴ걸요/EF, 려고요/EF, 데요/EF, 다는군요/EF, 느냐구요/EF, 더래요/EF, 구먼요/EF, ㄹ걸요/EF, ㄹ는지요/EF, 을걸요/EF, 애요/EF, 을게요/EF, 다요/EF, 냐고요/EF, ㅂ시요/EF, ㄴ다고요/EF, 는가요/EF, 던걸요/EF, 으세요/EF, 느냐고요/EF, 다지요/EF, 라면서요/EF, 라지요/EF, 읍니다요/EF, 다마다요/EF, 면서요/EF, ㄴ다요/EF, 디요/EF, 는대요/EF, ㄴ다지요/EF, 을라구요/EF, 라더군요/EF, 라나요/EF, 당게요/EF, 다는데요/EF, 다나요/EF, ㄹ라요/EF, ㄹ라구요/EF, 라서요/EF, 을래요/EF, 다더군요/EF, ㄴ다면서요/EF, ㄹ려구요/EF, 자구요/EF, 을께요/EF, 은가요/EF, 리요/EF, ㄴ다더군요/EF, ㄴ다는군요/EF, 던지요/EF, 느라구요/EF, 늬요/EF, 으리요/EF, 더라니요/EF, 러요/EF, 든지요/EF, 다던데요/EF, 씨요/EF, ㄴ다면요/EF, 서요/EF

4.2.6. 형태소 단위 vs 단어 단위

마지막으로 한국어 분석 말뭉치는 형태소 단위의 분석과 어휘 단위의 분석이 혼재되어 있다는 것을 이해할 필요가 있다. 우선 단어 형성법에서 합성어는 분석의 대상으로 삼지 않고 파생어만 대상으로 삼았으며, 파생어도 앞서 살핀 접사가 결합된 파생어만 대상으로 삼았다는 것을 이해해야 한다. 그렇지 않으면 자신이 연구 대상을 말뭉치에서 검색하지 못한 것을, 말뭉치에 없다고 하는 오류를 범할 수 있다.

5. 결론

이 논의는 사용하는 말뭉치의 성격과 특성에 대해서 잘 이해한 후, 말뭉치를 이용한 연구에 활용할 수 있어야 한다는 사실을 강조코자 한다. 따라서 말뭉치를 활용하여 연구를 진행할 때에는 기본적으로 다음과 같은 사항을 유의할 필요가 있겠다.

(1) 말뭉치 구축 시 발간된 보고서의 '지침' 확인 필요

(2) '검색'을 통해서 보고서의 나타나 있지 않은 '특징' 확인 필요

(3) 활용 말뭉치의 특성을 파악하기 위한 사용자 스스로의 노력

[참고문헌]

1.국립국어원(1997), 『21세기 세종계획』, 국립국어원.

2.국립국어원(1999), 『(21세기 세종계획)국어 기초자료 구축』, 국립국어원.

3.국립국어원(2000), 『(21세기 세종계획)국어 기초자료 구축 분과 특수자료구축 소분과』, 국립국어원.

4.국립국어원(2015), 『2015년 한국어 학습자 말뭉치 구축 지원 도구 개발 연구』, 국립국어원.

5.국립국어원(2015), 『2015년 한국어 학습자 말뭉치 기초 연구 및 구축 사업』, 국립국어원.

6.국립국어원(2016), 『2016년 한국어 학습자 말뭉치 기초 연구 및 구축 사업』, 국립국어원.

7.국립국어원(2017), 『2017년 한국어 학습자 말뭉치 기초 연구 및 구축 사업』, 국립국어원.

8.국립국어원(2018), 『2018년 한국어 학습자 말뭉치 기초 연구 및 구축 사업』, 국립국어원.

9.국립국어원(2019), 『2019년 한국어 학습자 말뭉치 기초 연구 및 구축 사업』, 국립국어원.

10.국립국어원(2019), 『형태 분석 말뭉치 구축』, 국립국어원.

11.국립국어원(2020), 『2020년 한국어 학습자 말뭉치 기초 연구 및 구축 사업』, 국립국어원.

12.국립국어원(2021), 『2021년 한국어 학습자 말뭉치 기초 연구 및 구축 사업』, 국립국어원.

13.김한샘·최정도(2020), 『연구용말뭉치구축의기초』, 경진(도서출판).

14.연규동·박진호·최운호(2003), 『인문학을위한컴퓨터』, 태학사.

15.홍윤표(2015), 『국어정보학』, 태학사.